INTERPLAY OF SOIL ORGANIC MATTER AND SOIL FERTILITY WITH STATE FACTORS AND SOIL PROPERTIES

H. JENNY [1], A.E. SALEM [2] and J.R. WALLIS [3]
University of California
Berkeley, Calif. - U.S.A.

INTRODUCTION

This paper is theoretical in that it attempts to combine conceptually and experimentally researches in soil fertility with those in soil genesis. To polarize the issue around personalities, DOKUCHAEV composed the first soil forming factor equation, LIEBIG assessed crop yields in terms of soil chemistry and plant nutrition, and HILGARD, standing between them, linked soil fertility problems to rocks, vegetation and climate. Our study is a modern hybrid of all three. It is based on a California collection of soil individuals that is highly stratified according to genetic factors and that was subjected to a large scale fertility experiment and was analysed chemically and mineralogically. The results suggest that interpretation of soil fertility studies is broadened and made more meaningful by integrating them with state factors (soil forming factors).

---

[1] Prof. Emeritus, reappointed.
[2] Postdoctoral Research Associate.
[3] I.B.M. Watson Research Center, Yorktown Heights, N.Y.

The Authors gratefully acknowledge the free use of the facilities of the University's Computer Center.

The Study Area, its State Factors and Soil Properties

The soils were collected to a depth of 20 cm in random fashion along a moisture traverse in California extending from the deserts of latitude 35° N to the high rainfall zone of latitude 41° N. This area had been previously sampled by Harradine and by Klemmedson [9] and was resampled by Jenny and Salem. The present study is essentially confined to Klemmedson's 97 soil samples on which additional work was performed.

As to *topography*, the exposures of the sampling sites were always south-east, hence invariant. The slopes (Sl) varied between 0 and 30%. All soils were well drained and free from influences of ground water tables.

Two kinds of *parent materials* (Pm) were selected. Acid igneous rocks (Ai), mostly granodiorites, and basic igneous rocks (Bi), mainly basalts and andesites. The two rock provinces were treated as dichotomous variables.

The *biotic factor*, the flora (Fl), was confined to *Pinus ponderosa* (Pi) and to « grass » (Gr) which included a variety of species of *Gramineae*, herbs and some desert shrubs. The forests had in part been selectively logged and the natural grasslands had been grazed, but neither had ever been cultivated.

The *climate* has hot, dry summers and cool, moist winters. Mean annual precipitation (P) of the sites ranged widely, from about 8 to 200 cm; mean annual temperature (T) was confined to 10-16°C.

The primary factors P and T were supplemented by potential evapotranspiration (Pet) and actual evapotranspiration (Aet) for which the available water holding capacity of all soils was assumed to be 15.24 cm (6.0 in.). Arkley's [1] leaching value (Li) was obtained by subtracting each mean monthly Pet from each mean monthly P. The 12 months summation of the positive values of P minus Pet is Li.

Although not usually considered state factors, *elevation* (El)

and *latitude* (Lat) were included because they might act climatically as seasonal modifiers of P and T.

The *age* t of the landscape is unknown, but the soils must be old as all collection sites were situated far beyond the area of possible disturbance by Pleistocene glaciers. The topography appears stabilized, man-made erosion is not noticeable and the profiles have the features of mature soils. Since nitrogen and carbon are believed to reach equilibrium in a few thousand years [7], we shall start with the assumption that soil organic matter is at steady state and hence time invariant.

On nearly all of the 97 soils 25 *properties* having pedogenic significance were determined by conventional methods. They are identified by symbols as follows: *C*, organic carbon, by dry combustion followed by CO_2 absorption, and corrected for free carbonates; *N*, total nitrogen according to KJELDAHL; *aci*, exchange acidity (me/100g) by leaching soil with $BaCl_2$ - triethanolamine at pH 8.1, according to MEHLICH. In the MEHLICH extract K and Na were determined by flame photometry, and *Ca* and *Mg* by atomic absorption spectrophotometry. In pastes pH was measured with glass electrode. Texture was analysed by pipette method as *clay*, *sa* (sand), and *si* (silt); *ME*, the moisture equivalent, by centrifuge method. Coarse and fine fraction are denoted by *cof* and *fin*, oven-dry moisture by *ov*.

Dr. I. BARSHAD [2] determined the clay minerals in per cent, specifically *fel* (feldspar), *cry* (crystobalite), *qtz* (quartz), *mia* (mica), *ver* (vermiculite), *mon* (montmorillonite), *kao* (kaolinite), *hal* (halloysite), *gib* (gibbsite) and *iro* (iron oxides).

Derived properties are *C/N*, *sba*, the sum of bases as Ca + Mg + K + Na, in me/100g, *cec*, the sum of cations (sba + aci), and *pac*, the percentage acidity given by 100 aci/cec. There are thus 29 soil properties.

On all 97 soils an elaborate pot test experiment was conducted. It is described in an ensuing section. The full treatment yields and the relative yields have the symbols FY and RY, respectively.

THE PEDOGENIC EQUATION AND ITS EVALUATION

Soil and the organisms in it and on it comprise the *soil ecosystem*. In this study the ecosystem properties pertain to certain aspects of vegetation but mainly to the 29 soil properties s_1, s_2, s_3 .... Among these the organic components N, C, C/N and their close relatives aci and pac are singled out for special emphasis.

The state factors [8] of the ecosystem, which determine its states, are its initial state L_o, here particularized as Pm, Sl, Fl, El and Lat, and the influx variables P_x — always measured in the environment — represented by P, T, Li, Pet and Aet. As mentioned, the age of the system (time factor) is tentatively left undetermined. The resulting, somewhat bulky sequence of equations is

$$(1) \qquad \mathrm{FY,\ RY} = f \begin{Bmatrix} \mathrm{N,\ C,\ C/N} \\ \mathrm{aci,\ pac} \\ s_1,\ s_2,\ s_3\ \dots \end{Bmatrix} = f \begin{Bmatrix} \mathrm{P,\ T,\ Li,\ Pet,\ Aet} \\ \mathrm{El,\ Lat,\ Pm,\ Sl,\ Fl} \end{Bmatrix}$$

Proceeding from right to left, our inquiry is directed to the relations among the state factors themselves, to the dependencies of soil properties on state factors, followed by relations among the soil properties themselves, and to the dependencies of crop yields on state factors and soil properties.

The evaluation of equation [1] relies on recent developments in multivariate statistical analysis which are geared to fast electronic computers [10, 11, 12, 13].

Variabilities of soil properties are characterized by variances. The dependency of one variable (Y) on another (X_1) is expressed by the least square regression equation

$$(2) \qquad Y = a + b_1 X_1$$

where a is the intercept and b_1 the slope coefficient. Individual values of Y or X_1 are designated as Y_i and X_{1i}, sometimes called scores. Means are denoted by $\bar{Y}$ and $\bar{X}_1$. The overall association of Y and X_1 is measured by the *correlation coefficient* r which is zero in absence of a trend and $+1$ or -1 for perfect collinearity.

For several predictor variables X_1, X_2, X_3, ... the linear multiple regression equation reads

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots. \tag{3}$$

The multiple correlation coefficient R is a measure of the success of regression. It lies between 0 and 1. The square of R or r is known as coefficient of discrimination R^2, r^2. It expresses the portion of the variability of Y that can be explained by the variations in X_1, X_2, X_3, ... Much of the ensuing discussion rests on the interpretation of R^2. As the number of X variables increases, R^2 tends to rise, but simultaneously degrees of freedom are lost, and if the gain in R^2 is small its adjusted value R_A^2 may actually decline. As we are not particularly interested in valid numerical predictions of Y from the X's but rather in the contributions of the individual X's in explaining the variance of Y, more emphasis will be placed on R^2 than R_A^2.

When the independent variables X_1, X_2, X_3 ... are highly self-correlated (collinear) the slope coefficients b become unstable, even meaningless as to sign. Regressing for example N against P (in.) and T (°F) gives

$$N = 0.350 + 0.0012\ P - 0.0055\ T \tag{4}$$

with $R^2 = 0.324$. Introducing Li (in.), which is highly collinear with P, results in

$$(5) \qquad N = 0.375 + 0.0037\ P - 0.0062\ T - 0.0029\ Li$$

The slope coefficient of P has tripled and that of Li is negative, which is absurd from the viewpoint of soil leaching. R^2 remains essentially unchanged as 0.327.

The handicap of self-correlation can be overcome by computing *principal components*. To illustrate, consider the correlation of aci and C of the 14 soil samples in Fig. 1, left. The solid, inclined line with angle θ (76.73°) passes through the center point which has $\overline{aci}$ equal to 6.656 me/100g and $\bar{C}$ equal to 2.133% as its coordinates. The line is the first principal axis (1P axis), constructed by minimizing the sum of the squares of the projections *normal* to this axis. The right-angle line through the center point is the second principal axis (2P axis).

Rotating the original coordinate axes through angle $(90-\theta)°$ to the unconventional right, for the sake of simplified pictorial presentation, produces the new coordinate system of Fig. 1, right. The new coordinates of the observation points are known as the first and second principal components (1PC, 2PC). For point P_i with aci of 12.64 me/100g (X_{1i}) and C of 4.11% (X_{2i}) the two component individuals (or scores) are

$$1PC_i = X_{1i} \sin\theta + X_{2i} \cos\theta = 13.25$$
$$2PC_i = X_{2i} \sin\theta - X_{1i} \cos\theta = 1.09$$

The 14 points give 14 individual values of 1PC and 2PC each. Their variances λ_1 and λ_2 (eigenvalues) are 10.485 for 1PC and 0.122 for 2PC. Since λ_1 is nearly equal to the ag-

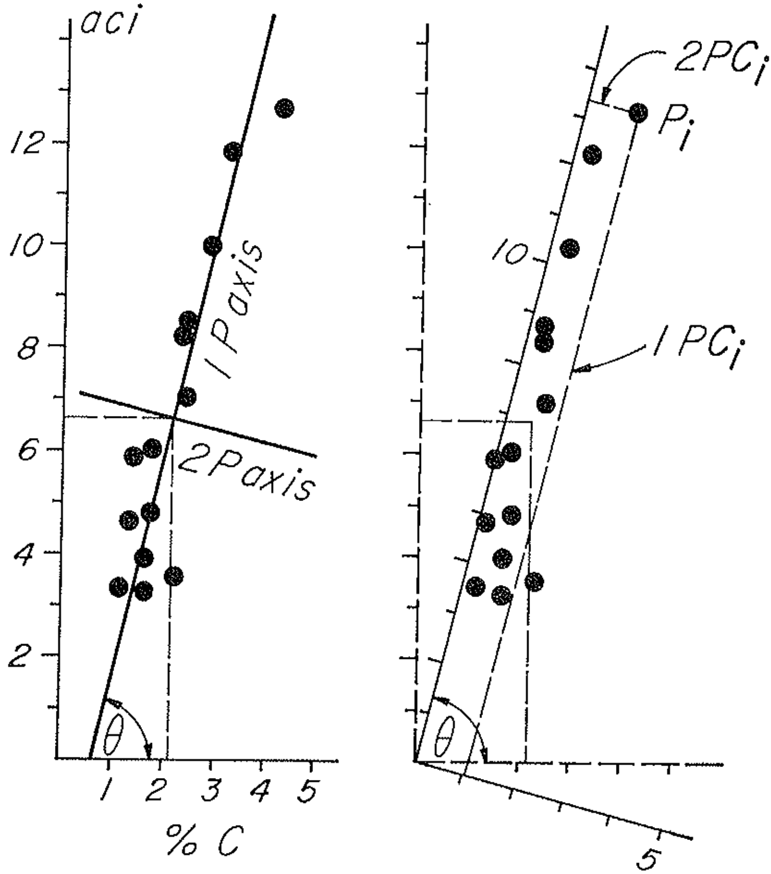

FIG. 1 — Illustration of converting original variables (aci, C) to first and second principal components (1.c. PC, 2PC).

gregate variances of aci (9.929) and C (0.669), the first principal component 1PC may be considered a numerical blend or composite of the original variables. It has no special name and may have ackward units. Parsimony has been achieved by replacing the two variables aci and C by one (1PC).

Three variables have three PC's and n variables have n PC's. Often one or two or very few component axes suffice to characterize an entire *cluster* of variables.

If the individuals of the components (the scores) are correlated with the individuals of the original variables the resulting r's are the component weights or loadings. Frequently, it is helpful to rotate the principal axes and arrive at easier identifications with the original soil properties. It was done in this paper by varimax transformation of the loading matrix. The new variables, here labelled PCV, may be used in place of Y and the X's in eq. (3).

The PCV regression technique gives for the aforementioned $N = f$ (P, T, Li) the equation

$$N = 0.346 + 0.0006\ P - 0.0054\ T + 0.0007\ Li \tag{6}$$

which is to be compared with eq. (5). The insertion of Li left R^2 unchanged (0.32). The slope coefficient of Li is now positive and about the same as that of P, as it should be. Moreover, the sum of the P and Li slopes in eq. (6) matches that of P in eq. (4), again as would be expected from the high collinearity of P and Li.

Correlation among State Factors

In Table 1 the correlation coefficients among P, T, Pm and Sl are very low. For practical purposes these factors are independent of each other. Importantly, the low r's guarantee that the distribution of Ai- and Bi-rocks has no climatic bias. As expected, temperature varies with El and Lat, and for the trivariate combination $T = f$ (El, Lat) the multiple correlation coefficient is high, $R = 0.892$. One could replace T by El and

Lat. From chemical and biological points of view T is the preferred variable.

Correlations of P and T with the derived climatic measures are either very low or very high. Surprising is the close association of P and Li ($r = 0.997$), the more so as P is merely an annual sum, whereas Li contains monthly combinations of climatic data. Also high are r for T versus Pet, and Aet versus log P. As the members of these pairs may be used here interchangeably, P and T are given preference. Pet, Aet and Li will be used as modifiers of P and T.

TABLE 1 — *Correlations among state factors. (For complete independence* $r = 0$, *for perfect association and « identity »* $r = \pm 1$).

		Primary state factors P	T	Pm	Sl
Primary state factors	P	1	−.20	−.08	.20
	T		1	.23	−.09
	Pm			1	.25
	Sl				1
Derived climatic variables	Pet	−.27	.98		
	Aet (1)	.84	−.03		
	Li	.997	−.24		

(1) For Aet and log P, $r = 0.97$.

Plant life deserves special comment as it occurs on both sides of the state factor equation; on the right as flora, that is, the pool of species as genotypes, on the left as vegetation, the species as phenotypes. As judged from the configuration of the landscape the seeds of pines and grasses may reach any site of the transect. For this reason the flora chosen is the invariant biotic factor.

The correlation coefficient of vegetation with P amounts to 0.728 for all soils. It expresses the predominance of pines over grasslands in the humid portion of the transect.

Dependence of Soil Properties on State Factors

Before attempting to solve the general regression equation in which all state factors are variables, it is pertinent to assess the dichotomous factors parent material (Pm) and flora (Fl).

Since *parent material* is neither correlated with P, T and Sl (as in Table 1), nor with Fl ($r = -0.139$), the Ai and Bi soils may be compared directly, as in Table 2. The texture fractions sand, silt and clay are essentially uncorrelated with state factors (e.g. all r's for clays are less than 0.229) and, therefore, the marked texture variation between Ai- and Bi-soils may be considered a direct reflection of the differences in Ai- and Bi-rocks. In fact, regressing a soil property with sand and clay equals that with Pm, though not entirely so. The sum of bases and the suite of clay minerals likewise are characteristically different, as are many other properties, but, since these also vary with climate they lack a one to one correspondance with the two rock types.

Since the distribution of *pine forests* and *grasslands* is strongly P-dependent, pine soils and grass soils cannot be compared directly, save in the arid-humid transition zone (56-76 cm) where both occur in the same climate as random patches of pines surrounded by grassy areas.

TABLE 2 — *Mean values of properties of soils derived from acid igneous (Ai) and basic igneous (Bi) rocks. (All Ai - Bi differences are significantly different at the 1% level or less.)*

Soil properties	Ai-soils	Bi-soils
clay, %	11.6	21.2
silt, %	21.2	33.0
sand, %	58.0	34.5
C %	1.74	2.80
N %	0.074	0.121
bases (sba) me/100g	5.33	10.86
1 PC of clay minerals	ver mia qtz mon	mont mia gib hal

The data in Table 3 include KLEMMEDSON'S samples and additional ones collected by the authors. Compared to pine soils, the grasslands have low C/N ratios, 12-13 versus 25-28, and this highly significant differentiation is brought about by C rather than N. The r-value for C/N with Fl is high, 0.809. Parent material does not modify C/N, confirming the results in Table 2. Carbon and aci are much higher in pine soils than in grass soils, at the 1% level of significance for Ai- and Bi-soils together. As in Table 2, the Bi-soils have a high base (sba) status, but vegetation does not cause a differentiation at the 10% level of significance.

TABLE 3 — *Effect of vegetation and parent material on means of soil properties in the arid-humid transition zone* ($P = 23\text{-}30'' = 58\text{-}76$ *cm*).

	Grasslands		Pine Forests		Grass with isolated pines
	Ai-soils $n = 10$ (1)	Bi-soils $n = 4$	Ai-soils $n = 8$	Bi-soils $n = 5$	Ai-soils $n = 6$
P	25.9	26.0	28.4	25.2	25.3
C/N	12.7	12.5	25.8	27.9	14.6
aci	2.55	4.45	8.77	13.00	2.49
C	1.03	1.31	2.98	3.20	1.11
sba	6.28	16.91	7.19	18.27	7.63

(1) n = number of samples.

The last column in Table 3 has pertinent eco-pedological significance. It describes the situation under isolated, 100-200 year old pine trees surrounded by grassland. Intrinsically, the soils are grassland soils with a slight displacement toward pine soils, which, however, is statistically of low significance. The lesson is conveyed that a single pine tree does not make a pine soil, rather a permanent cluster of pine trees is needed that imprints its marks throughout centuries and millennia. The soil criteria may have diagnostic significance for deciphering vegetational history.

To illustrate the *regression procedure* with dichotomous variables consider the 40 Bi-soils in Fig. 2. The dashed line re-

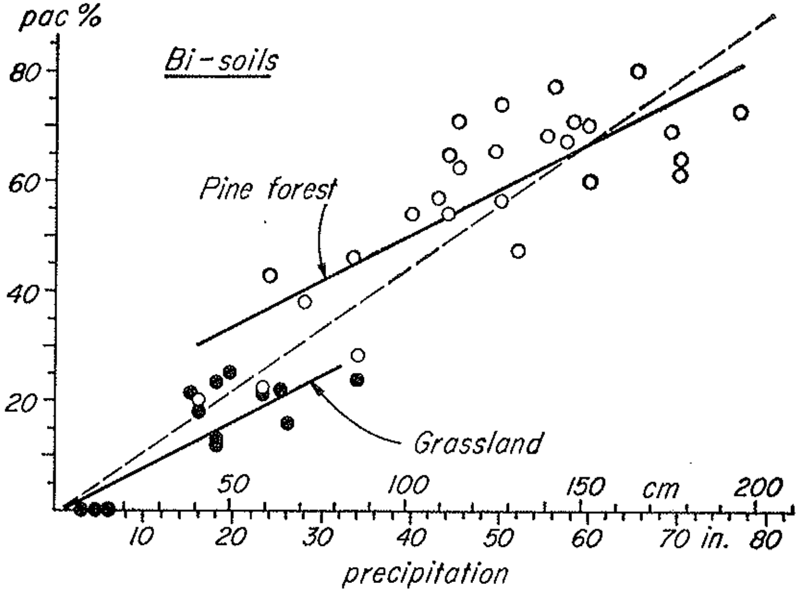

FIG. 2 — Dependency of percentage acidity (pac) on mean annual precipitation and vegetation groups.

lates aci to P, regardless of the kind of vegetation. The equation is

$$\text{pac} = 0.592 + 1.128\ P$$

with $r^2 = 0.834$ and $r = 0.913$. This r^2 registers the direct effect of P on aci and includes the indirect effect of vegetation which is correlated to P with a r value of 0.736.

Introducing the flora (Fl) as pac $- f$ (P, Fl), its value for pines is set as 1.000 and for grassland as -2.0769, which is the ratio of the 27 pine soils to the 13 grass soils to give a mean value for $\text{Fl} = 0$. The equation is

$$\text{pac} = 10.997 + 0.841\ P + 5.582\ \text{Fl}$$

[1] *Jenny* - pag. 13

with an improved R^2 of 0.883 or $R = 0.940$. This new R^2 expresses the combined influence of P and Fl. Under favorable circumstances the difference $0.883 - 0.834 = 0.049$ may be assigned to Fl directly. Replacing Fl by the pine and grassland value, brings forth the two equations

$$\text{pines:} \quad \text{pac} = 16.579 + 0.841\ P$$
$$\text{grassland:} \quad \text{pac} = -0.596 + 0.841\ P$$

Recognizing that pine and grass curves could have slightly different slopes, we formulate

$$\text{pac} = 10.964 + 0.848\ P + 5.827\ Fl - 0.0115\ Fl\ P$$

There is no significant improvement in R^2. The pine and grassland equations are

$$\text{pine:} \quad \text{pac} = 16.790 + 0.834\ P$$
$$\text{grassland:} \quad \text{pac} = -1.138 + 0.872\ P$$

The dependency of the organic soil constituents N, C, aci, pac on the state factors, summarized in Table 4, was evaluated in the following manner. Choosing N as an example, it was first regressed with P, then with P + T, then with P + T + Pm, and so forth. Within each of the first six constellations listed in Table 4 the individual R^2 contributions were nearly invariant and the slope coefficients of the factors remained fairly stable, hence the state factors could be assessed separately by conventional regression.

The four organic criteria N, C, aci, paci behave highly individualistically. They are discussed under the following headings:

Climate (P, T). Nitrogen is least affected by rainfall; only 22.8% of the entire N variance is explained by P. Percentage

TABLE 4 — *Variability of soil properties assigned to state factors. (% R^2 values obtained by stepwise conventional regression).*

		N	C	aci	pac	C/N (1)
1	P	22.8	45.2	45.5	80.9	20
2	T	5.4	9.5	3.6	0.3	2
3	Pm	16.3	6.0	17.4	2.0	1
4	Sl	1.2	1.3	3.5	0.3	6
5	Fl	0.2	3.9	4.4	8.9	40
6	Lat	5.9	3.1	1.9	0.4	7
	Sum	51.8	69.0	76.3	92.8	77
	R	0.720	0.831	0.874	0.963	0.878
7	Min	63.4	74.2	79.7	2.6	80.7
8	cl	5.2	1.3	0.7	0.2	1.6
	Total sum	68.6	75.5	80.4	95.6	82.3
	R	0.823	0.870	0.897	0.978	0.907

(1) Computed by PCV regression.

acidity (pac) is at the other extreme, with $R^2 = 80.9\%$. Carbon and aci occupy intermediate positions. For N and C relating to T, r is negative, as had been observed in many previous investigations.

Parent material (Pm). Nitrogen and aci react strongly to variations in parent material, C much less so, pac hardly at all.

Parent material exerts its role mainly through soil texture and clay mineralogy.

Slope (Sl). Owing to careful topographic controls during field sampling, the slope effect is small, except for C/N.

Flora (Fl). The role of species is highly specific, being near zero for N and 8.9% for percentage acidity (pac). It would seem that biologic N fixation which in the long run controls the steady state of N is but slightly affected by vegetation, whereas C, whose source is photosynthesis, is much more so. This is corroborated by the C/N ratio to be discussed presently.

Latitude (Lat). Carbon and particularly N have significant negative coefficients with latitude. It implies that for any chosen pair of P and T coordinates soil organic matter increases toward lower latitudes. The reason is obscure. Lat may mitigate mean annual P and T, or register systematic biotic variation. At any rate, the trend is in accord with observations made in India and Latin America and leads to the prediction that the soils having the highest climatically controlled humus content should be found at the equator at high elevations (e.g. Mt. Kilimanjaro).

The sums of the columns of the six factor allocations divulge variance explanations of 52% for N and a spectacular 92.8% for pac. The row of the conventional R values indicates — exclusive of N — that all multiple correlation are highly successful.

The *C/N ratio* could not be regressed conventionally because high interaction and collinearity forfeited the consistency of the individual factors. Thus, C/N correlated with P alone to the extent of $R^2 = 44.7\%$, but when Fl was inserted it reduced the P share to only 0.4% and amassed for itself 73% units of the total R^2 sum of 77.0%. The C/N column was obtained by PCV regression which successfully masters the interaction problem. Precipitation explains 20% of the va-

riance and this comprises both the direct effect of P on C/N and the climatic control of vegetation. The species effect on C/N, as flora Fl, adds 40% to R^2.

Let us divert for a moment from organic matter and look at the regimen of clay minerals. It responds to all state factors. Montmorillonite prevails at low rainfall [2], for its correlation with P is negative, with r equal -0.534. Kaolinate and halloysite are negatively correlated with each other, $r = -0.800$, and its IPC regressed with P and T gives $R^2 = 11.3\%$ which rises to 61.4% upon the addition of Pm. The first PC of all clays taken together is affected by climate to the extent of $R^2 = 31.9\%$ and by all factors as $R^2 = 82.5\%$ a remarkably high value for an inert mineral soil constituent.

Granting that soil organic matter is vastly more dynamic than clay alteration, we may be justified in considering the clay matrix as a stable physical and chemical frame work for biochemical reactions, a sort of parent material for the organic substances. Accordingly, the clays are introduced as factor No. 7 in the regression chain of Table 4.

Owing to the dependency of clay minerals on climate it is not possible to isolate their share in the regression coefficient. Only the total contribution is listed in Table 4. Nitrogen reacts most strongly, the rise in R^2 being 11.6%. The others have their R^2's augmented by 3-5%.

Derived climatic variables. If Li, Pet and Aet are added for completeness sake, also El, a further slight increase in R^2 is produced for N, but barely for the other variables.

Looking at the *total sum of R^2* in Table 4, the factors explain 68.6% of the variance of N and an amazing 95.6% ($R_A^2 = 94.4\%$) for pac. C/N is also very high, followed by aci and C.

Clusters of highly correlated soil properties, such as the acidity dyad pH*pac or the organic matter triad N*C*C/N — where the stars indicate that the properties have been cor-

related — may likewise be coordinated with state factors. It results in economy of functions. One function, with the first principal component of the cluster as the dependent variable, takes the place of two or three. High R^2 values may be achieved (Table 5) but not quite as high as the highest R^2 of the top ranking partner in a cluster.

TABLE 5 — *Ordination of first principal components of dyads and triads with state factors. R^2 values (%); all soils.*

Cluster (1)	Climate alone (P, T)	All State Factors
N*C	44.6	64.1
pH*pac	76.8	91.2
N*C*C/N	49.0	74.0
N*C*aci	48.9	74.4

(1) The stars indicate that the properties are closely correlated.

Ordination of soil properties confined *within Ai-* and *Bi-rock provinces* was carried out with PCV technique. The R^2 values for the constellation P, T, Sl, Fl and Lat are reported in Table 6. They are high, and are even higher if texture is included as a criterion of composition variability within the Ai- and Bi-rocks. As expected, the ranking of the clay minerals differs in the two materials.

Not shown are the standardized slope coefficients which record the thrust of the individual state factors. Among the Ai-soils the flora contribution to R^2 is 3% for N, 26% for C

TABLE 6 — *Variance of selected properties of Ai- and Bi-soils explained as R^2 by the state factors P, T, Sl, Fl and Lat, according to the PCV method. (The R^2's in parenthesis express the inclusion of soil texture (sand, clay) as a state factor to characterize the mineralogical variability within the Ai- and Bi-rocks).*

Ai-soils		Bi-soils	
Soil property	R^2	Soil property	R^2
pac	94 (95)	pac	89 (93)
C/N	82 (83)	sba	77 (78)
pH	80 (82)	aci	74 (75)
aci	68 (85)	C/N	73 (73)
C	67 (78)	C	69 (69)
mica	66 (73)	ME	66 (77)
verm	65 (65)	pH	65 (66)
sba	35 (67)	mica	64 (67)
ME	28 (84)	mon	60 (67)
N	25 (45)	N	53 (57)

and 43% for C/N. At a given climate Fl also modifies the proportions of the clay minerals mica, kao, ver and mon. Among the Bi-soils the flora contributions to total R^2 are 16% for mon, 27% for hal and 38% for kao.

All these R^2 values furnish a beautiful numerical confirmation of the ideas of HILGARD and especially of Dokuchaev's equation (4). We believe that the data in Tables 4, 5 and 6 constitute its first complete solution.

Our findings are seemingly in conflict with a recent trend of thought that denies the validity of soil forming factor equa-

tions. It is contended that most soils are paleosoils or fossil soils that owe their existence to previous climates. It cannot be overlooked, however, that the high correlation in Tables 4, 5 and 6 pertain to present-day climate, slopes, vegetation, etc. Either there has not been a pedologically measurable change in climate in the study area, or the soil features assayed here readjusted themselves to modern climate. In this light the expression paleosoil is ambiguous without the explicit specification of the kind of properties one has in mind. Evidentely, organic soil constituents reflect todays state factors.

Interaction of Nitrogen and Carbon with Soil Properties

Comparing each of n soil properties with every other produces $n(n\text{-}1)/2$ *correlation coefficients*, or for our collection 406 r's. If classified in intervals of 0.200 units, the frequency distribution in Table 7 results. Very low associations of pro-

TABLE 7 — *Frequency distributions of* 406 *correlation coefficients* $|r|$ *among* 29 *soil properties.*

Ranges of r	Number of r values	Examples (1) of correlating properties with emphasis on N, C, C/N, aci, pac
0.999-0.800	13	N*aci, sba*Ca, pac*pH, C*N
0.799-0.600	28	Ca*Mg, sa*clay, hal*ME, C/N*pH
0.599-0.400	76	C*pH, N*iro, mi*ver, C*mi
0.399-0.200	113	mon*Na, si*aci, C*cry, clay*pac
$<$0.199	176	C/N*kao, N*K, Ca*aci, C*sba

(1) The star denotes that the properties chosen are correlated.

[1] *Jenny* - pag. 20

perties dominate, but it is the 40 high ones that are important. As judged from the arbitrarily selected examples, the organic constituents occur in all correlation ranges.

In the large overall correlation matrix (not shown) the 84 r-values which involve N, C, N/C appear erratic and puzzling. Why should N be highly covariant with aci ($r=0.852$) but barely so with K ($r=0.006$)? Or, to consider clays, why should C be positively correlated with halloysite and negatively with montmorillonite? Such questions are basic to pedogenesis and they figure prominently in discussions on soil classification. Answers are usually sought in terms of physical, chemical and biological soil processes, but these are still too poorly understood to offer much help. The entire problem of correlations of soil properties is formidable and largely unsolved.

Light is shed on the multitude of r-configurations of N, C, C/N by examining the correlation coefficients of the soil properties with their climatic state factors. It becomes at once apparent that C and N have positive r values with all those soil properties that are positively correlated with precipitation and negative r values with those that are negatively correlated with P. For carbon the design is displayed in Fig. 3. On the horizontal axis are plotted the $\pm r$'s for $s=f(P)$, on the vertical the $\pm r$'s for $C=f(s)$, where s indicates any soil property. As seen in the first quadrant, aci is well correlated with P($r=0.675$) and, in turn, C is strongly collinear with aci ($r=0.945$). In the third quadrant r for $pH=f(P)$ is markedly negative ($r=-0.772$) and so is $C=f(pH)$, with $r=-0.574$.

Clearly, it is the impact of precipitation on all soil properties that regulates their correlations with carbon. This observation strongly suggests that interpreting mutual associations of soil properties is facilitated by invoking their bondage with the state factors.

High correlation coefficients, such as $r=0.945$ for C*aci, are indicative of the existence of stable dyads, triads and

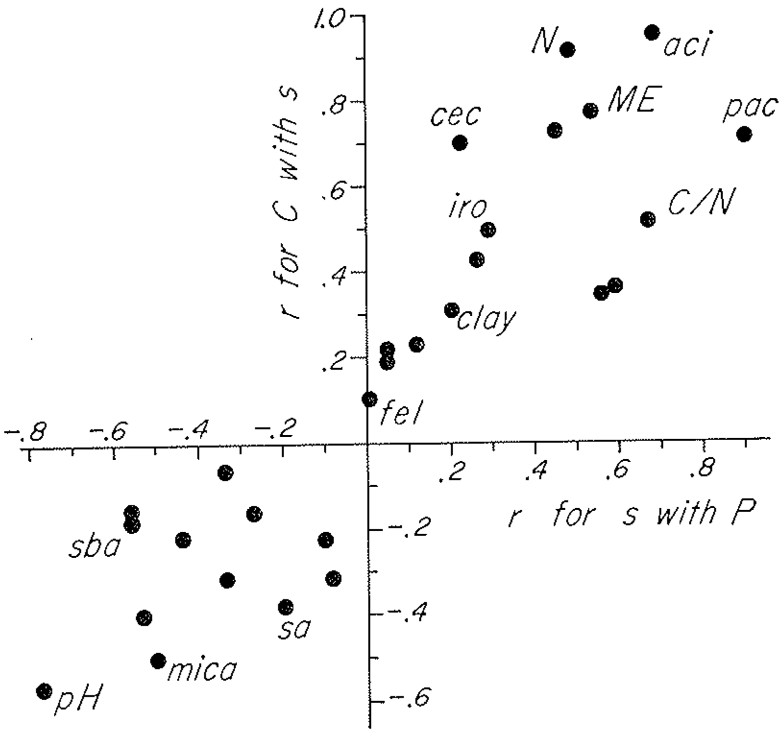

FIG. 3 — A correlation between correlation coefficients: r for $C = f(s)$ against r for $s = f(P)$, showing that carbon correlates highly with those soil properties that are highly correlated with precipitation (P).

clusters of properties. For the dyad C*N with r equal to 0.908, any covariance of C with another property s is mirrored in the r value of N with that same s. Thus, for $C = f(\text{cec})$, r is 0.701 and for $N = f$ (cec) r is 0.765. N*C*aci and sba*Ca*Mg are triads. A 6-variable cluster is given by N*C*aci*cec* ov*ME, but it has not been studied.

As said earlier partnership of variables may be coded as

[1] *Jenny* - pag. 22

principal components, especially if the first component (1PC) has a high eigenvalue λ_1. Thus λ_1 of N*C*aci accounts for 93.5% of the combined variances of the three. Its 1PC was regressed with state factors in Table 5. Correlating the component individuals (scores) of N*C, sa*si, paci*pH, N*C*aci, and of the ten clay minerals with those of P and C produces r's that fit into the scatter diagram of Fig. 3.

Recent advances in cluster research, Factor Analysis — used in soil fertility studies long ago by FERRARI et al. [5] — and numerical taxonomy may stimulate attempts at ordinating soil properties. Fig. 3 provides a convincing argument that state factors deserve to be included in the ordination. In fact, there is no compelling reason, philosophical or otherwise, why they should not.

The question how much *time* it takes nature to create a high correlation among soil properties is answered by the mud flows of Mt. Shasta (P = 123 cm, T = 9.6°C) in California. The sequence of flows, originally established by DICKSON and CROCKER [3], is sketched schematically in Fig. 4. The senior writer reexamined the field stratigraphy and confirmed the essential overlays, especially that of D on E, but could not be sure whether or not B different from C. Moreover, C-14 dating set the age of B as 300 years, not 60, as the authors had assumed. At a slightly higher elevation an older flow F was found but its exact stratigraphy has not yet been worked out. It is not shown in the flow sketch in Fig. 4.

In these flows 12 random samples (0-20 cm) were collected and analysed for aci and C. The regressions together with the r values are shown in Fig. 4. Surprisingly high coefficients are reached in a few decades and centuries even though neither C nor aci appear to have individually achieved steady state conditions. The age of the older flows has not yet been established but, as judged by inorganic and textural profile features, it cannot exceed a few thousand years.

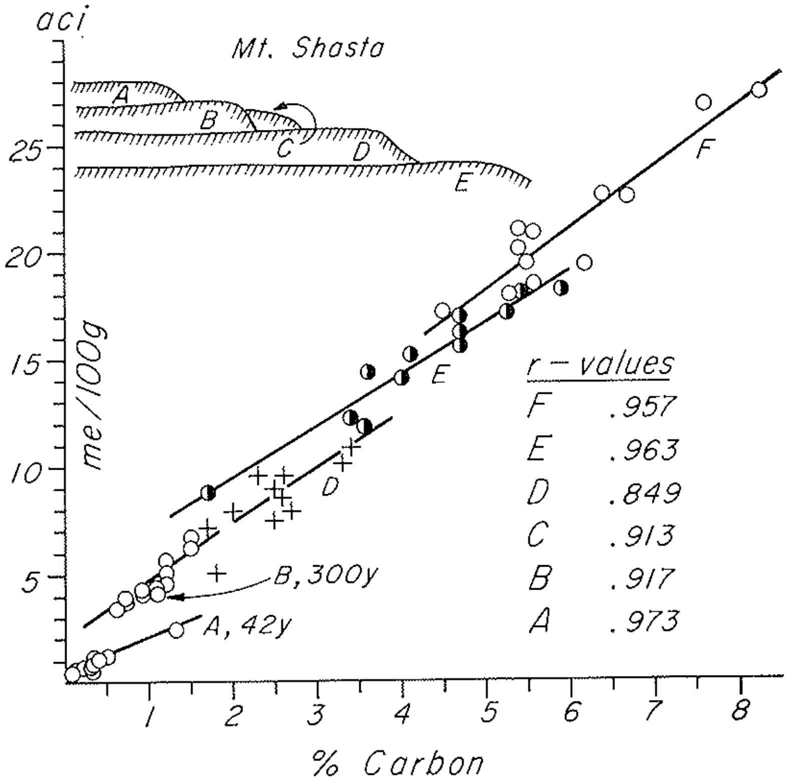

FIG. 4 — Effect of age of soil on the correlation coefficient of aci versus C. Mud flows of Mt. Shasta, dated as 42 years (A), 300 years (B), and a few thousand years for D, E and F.

The average number of me. of acidity per gram of carbon is close to 3.5, or 42 me acid per mol of carbon. For the corresponding Bi-soils (pines) the mean value is 5.0 me/gram C. The difference of 1.5 me of acidity may be attributed to the clay fraction which is high in Bi-soils but practically nil in the Shasta mud flows.

[1] *Jenny* - pag. 24

Mineralization of Soil Nitrogen

For a fertility test [9] on the Berkeley campus quantities of the 97 soils were placed in pots, four times replicated. The 2000 pots, all randomized, were processed simultaneously. Barley was used as an indicator plant. The « full treatment » consisted of an application of ammonium nitrate, monocalcium phosphate and sodium sulfate, designated as NPS. Potash was not included because previous experience suggested absence of K deficiencies. One of the partial treatments, indexed as PS, had nitrogen omitted. The *relative yield* RY in per cent was calculated as

$$RY = \frac{\text{mean yield of PS}}{\text{mean yield of NPS}} \cdot 100 \; .$$

It measures the soil's supplying power of nitrogen to the barley plants under the conditions of the experiment. The mean of all RY's was 19.6% with extremes of 6.0% and 72.0%. The variance of the RY population was large, $123.0(\%)^2$. As the RY yield is essentially an expression of the rate of mineralization of organic soil nitrogen, the wide yield variations testify to highly variegated soil microbiological activities. These will be assessed by way of state factors and soil properties.
will be assessed by way of state factors and soil properties.

Yields assigned to state factors. If soil properties are functions of state factors and if barley yields are controlled by soil properties, then the yields observed at Berkeley should exhibit a linkage to the state factors at soil origin, in accordance with eq. (1).

For all soils taken together the correlation coefficient of RY with mean annual precipitation (P) is 0.595. If, in addition, T, Pm, Sl, Fl and Lat are placed in multiple regression, R

becomes as large as 0.714. Limiting regression analysis to Ai soils exclusively, the PCV method rises R to 0.80.

Even more striking is the behaviour of the yields FY of the full treatment pots NPS. Plant growth was expected to be uniformly good, but it was not. The yields of the Bi-soils at Berkeley declined systematically in proportion to the mean annual precipitation at the site of soil origin, with r equal -0.704. The impairment could be documented as phosphate fixation caused, presumably, by the demonstrable enrichment of gibbsite clays and iron oxides with rising rainfall. Multiple regression of FY with P, T, Sl, Fl and Pm (as texture) gives $R=0.816$, and, if the clay minerals are included, R becomes 0.864.

There is good evidence, then, that soil fertility measured as RY or FY at Berkeley still bears the imprint of the factors of soil genesis, even though the soils had been moved, disturbed and manipulated.

Yet, a R^2 of 67% is still one third below the maximum figure of 100%. The source of the discrepancy most likely resides in the factor parent material which is poorly identified. No appraisal was made of chemical constitution, including the plant-important microelements, or of the mineralogical design. Since many of the rock components are transfered to the soil and as soil properties diversify microbial activity and plant growth, yield correlations should be higher for soil properties than for the state factors. But how much higher?

Genetic soil properties as yield determinants. If one were to sample soils haphazardly in a « random-landscape » — a hypothetical construct in which all state factors are geographically randomized — the probability of observing a close accord of yield and soil would be small because the soil properties are likely to be uncorrelated. The soil collection would be highly heterogeneous.

Contrarily, if an assemblage of soils is highly stratified according to continuous pedogenic sequences it is monotonely

homogeneous. Yield functions should be promising, for all properties are arranged in orderly sequence and are intimately interrelated, viz. collinear.

The present collection which comprises calcareous, neutral and acid soils, sands and clays, humus-rich and humus-poor soils, gray desert, non-calcic brown and lateritic soils would suggest extreme heterogeneity were it not for the fact that the sites were chosen to yield gradual transitions, safe for the discontinuities of Ai- and Bi-rocks and of grass and pine flora. Accordingly, a certain degree of yield prediction is to be expected, especially when the samples are limited to within the discontinuous factor sets.

Correlation of yields with organic constituents are consistently high (Table 8). Either N or C alone, or in combination, or as principal components account for 63-64% of the yield variance. A completely randomized soil collection would not generate such high R^2 values.

TABLE 8 — *Correlation coefficients* r, *R*, *of relative yields with organic soil constituents and their principal components (PC). All soils.*

Soil variables	r or R	Soil variables		r or R
N	0.791	1PC(N*C) and C/N		0.802
C	0.800	1PC	of N*C*C/N	0.780
N, C	0.789	1, 2PC		0.787
N, C, C/N	0.802	1, 2, 3PC [1]		0.791

[1] Linear regression of RY with first, second and third principal components of the cluster N*C*C/N.

The pertinent contract between RY and total N is plotted in Fig. 5. The regression lines have the following forms:

all soils: $RY = 3.22 + 175.98\ N\ (r = 0.791)$

Ai-soils: $RY = 5.13 + 166.86\ N\ (r - 0.693)$

Bi-soils: $RY = 0.91 + 197.58\ N\ (r = 0.833)$

The scatter of the points, which is more erratic for Ai than Bi, is to be attributed in large part to those soil properties that govern the rate of nitrification and mineral uptake by plants. What are these properties? Soil fertility specialists would analyse for nitrates, including an incubation test, available K and phosphates, micronutrients, maybe nitrites and soluble aluminum, all of them properties that are not contained in our set of 29. Yet, *if collinearity among soil properties is high, the choice of variables should not be crucial,* and even the genetic characteristics should function as predictors of yields.

The R^2 values in Table 9, obtained by stepwise regression analysis tell a revealing story. For all soils (Ai + Bi) together, nitrogen alone explains 63% of the yield variability. Four additional soil properties, selected by the computer, augment R^2 to 71%. Ten statistically chosen properties raise R^2 to 73%. Regression with the PCV variables gave R^2 of 89%, a high value. Its standardized slope coefficients were positive, viz. yield enhancing, for N, C, hal, kao, clay, gib, ver, in that order, and negative, viz. yield retarding, for aci, pac, Mg, pH, iro, again in that order.

Confining the inquiry to one kind of parent material, either Ai or Bi, is equivalent to resolving the sample collection into two subsets which may or may not have different population parameters. Compared to the whole collection, multivariate regression with ten soil properties augments R^2 within the two rock subsets, from 73% to 81-82%. With further sample

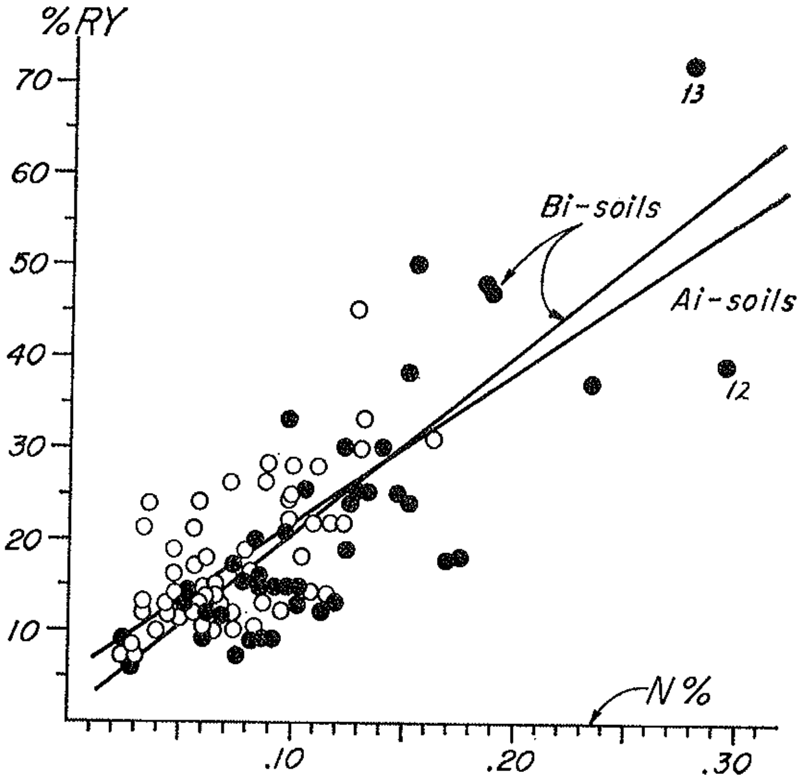

FIG. 5 — Relative yield RY plotted against total nitrogen content of soil. RY is an index of rate of mineralization of organic N.

modulation (pine soils only) R^2 rises still higher (87-89%), even though mean annual precipitation ranged widely, between 16 and 77 in. (40-196 cm). The unexplained variance is of the order of experimental error in pot test technique. Modulating the state factors evidently improves the correlation between yields and soil properties.

The scatter of the points in Fig. 5 is meaningfully explained. Thus for the Bi-pine soils No. 12 and 13, the former has an

TABLE 9 — *Explaining yield (RY) variance as R^2(%) by modulating state factors according to parent material and flora.*

Number of soil properties including N / State factor modulation	% R^2		
	N = 1	N + 4 = 5	N + 9 = 10
Ai -and Bi-soils together . .	63	71	73
Ai-soils only	48	72	81
Bi-soils only	69	75	82
Ai-soils, pines only	47	80	87
Bi-soils, pines only	66	84	89
Bi-soils, pines only, mult. regr. coeff. R	0.81	0.92	0.95

observed yield of 39%, the latter of 72%. The calculated yields are 38% and 73% respectively. Soil 13 is more fertile than soil 12 because it has much less acidity (aci) and much more Ca.

The middle column headed N + 4 = 5 deals with regressions with those 5 soil properties that are at the top of the list of the 29, as far as their contributions to R^2 is concerned. For all state factor subsets (rows) the five invariably include N and aci; N/C and clay occur 3 times each and C, pH and pac twice each. Neither their order nor signs are consistent from one subset to another, nor should consistency be expected between discontinuous fields.

It is indeed astonishing that 5 properties should explain over 80% of the yield variation when the plant itself needs over a dozen different elements, not to mention suitable micro-

soil structure for microbial activities. The explanation is to be sought in the high collinearity of many soil properties brought about by the systematic sampling procedure according to state factors.

SUMMARY

Theoretically, crop yields, soil properties and factors of soil genesis (state factors) are interrelated as a multivariate system.

To test this broad concept, 97 soils were collected in California along a moisture transect reaching from the desert to the perhumid region. Parent rock, exposure, slope and natural vegetation (pines, grasslands) were rigidly controlled. The mathematical model was based on linear regression analysis involving observational variables as well as principal components with and without varimax rotation.

The variability of soil properties, especially those related to organic matter, could be accounted for in high degree by the present-day soil forming factors, especially climate and natural vegetation. The old idea that N and C are linked by common microbial genesis could not be confirmed as the two elements behaved differentially. Nitrogen was but moderately related to climate and barely so to flora, but strongly to parent material and clay mineralogy. Carbon, on the other hand, responded strongly to precipitation and temperature and markedly to plant species, which was especially expressive in the behavior of C/N.

Clay mineralogy correlated with climate and parent material, and, at a given climate, also with floristic differentiation.

Correlation coefficients (r) among the soil properties ranged between near-zero and near the maximum of 1.00. The coefficients of carbon with the other soil properties were governed by their dependencies on precipitation. It takes but a few

decades or centuries to establish high r values between such pairs as carbon and acidity.

Soil fertility was ascertained in a large-scale standardized pot test experiment. The yields variations still reflected the state factors of soil origin. Degree of correlation of yields with multiple soil properties tended to depend on the genetic ordination of soil samples; the more heterogeneous the soil group as to dichotomous factors such as rocks and flora, the less was the multiple coefficient of correlation.

REFERENCES

[1] ARKLEY R.J., *Calculation of carbonate and water movement in soil from climatic data.* « Soil Sci. », *96*, 239-248 (1963).

[2] BARSHAD I., *The effect of a variation in precipitation on the nature of clay mineral formation in soils from acid and basic igneous rocks.* « Int. Clay Conf. Proc. », *1*, 167-173 (1966).

[3] DICKSON B.A. and R.L. CROCKER, *A chronosequence of soils and vegetation near Mt. Shasta, California.* « Journ. Soil Sci. », *4*, 123-141, 173-191 (1953-1954).

[4] DOKUCHAEV V.V., Writings. Akademia Nauk, Moscow, *6*, 381 (1951).

[5] FERRARI TH. J., H. PIJL and J.T.N. VENEKAMP, *Factor analysis in agricultural research.* « Neth. J. Agric. Sci. », *5*, 211-221 (1957).

[6] JENNY H., *Comparison of soil nitrogen and carbon in tropical and temperate regions.* « Missouri Agr. Exp. Sta. Res. Bul. », *765*, 1-30 (1961).

[7] — *Bodenstickstoff und seine Abhängigkeit von Zustandsfaktoren* « Zeitschr. Pflanzenernähr., Düng., Bodenk. », *109*, 97-112 (1965).

[8] — *Derivation of state factor equations of soils and ecosystems.* « Soil Sci. Soc. Amer. Proc. », *25*, 385-388 (1961).

[9] KLEMMEDSON J.O. and H. JENNY, *Nitrogen availability in California soils in relation to precipitation and parent material.* « Soil Sci. », *102*, 215-222 (1966).

[10] MORRISON D.F., *Multivariate statistical methods.* McGraw-Hill Book Company, New York, pp. 338 (1967).

[11] SEAL H., *Multivariate statistical analysis for biologists.* Methuen Company, London, pp. 209 (1966).

[12] WALLIS J.R., *Multivariate statistical methods in hydrology - a comparison using data of known functional relationship.* « Water Resources Research », *1*, 447-461 (1965).

[13] — *Factor analysis in hydrology - an agnostic view.* « Water Resources Research ». In print. (1968).

DISCUSSION

Chairman: S. A. WAKSMAN

WAKSMAN

We have, according to our programme, fifteen minutes for discussion of this most interesting paper of Dr. JENNY. I would like to take advantage of the fact that I am the Chairman in order to suggest that there is one other, very important factor that might be considered. Dr. JENNY mentioned the kinetics of the soil, which is extremely important. The microbiological population of the soil is not necessarily the number of organisms or the kinds of organisms but the total activity of the population, as measured either as nitrogen liberation or carbon dioxide liberation, or by some other measurement. I think this might add to the factors which HILLGARD took into consideration. What do you think?

JENNY

This touches upon an important question, namely, the relationship and nature of the variables. In what aspect are the microbes in this area an independent variable? The idea is that every soil of the whole area during thousands of years received through wind and blowing and other ways all the microbes of the area, and then the activity of the microbes, which eventually brings about the

steady state of the soil, is a dependent variable and is really controlled by these same state factors. Microbial activity is a process, and processes are not part of the state factor approach.

DHAR

What was the calcium phosphate status of your soil? Did you take up that question? Calcium phosphate status of a soil is of vital importance to fertility.

JENNY

In some of the fertility experiments we added phosphorus to bring about a modification of fertility and that is the way we learned that phosphate fixation was a function of the climate of these soils. And likewise, the calcium content was altered. But we have not introduced the natural phosphate level into the calculations of the factors' correlations. It would have been better if we had done this, but we didn't.

SWABY

Dr. JENNY, I notice that on most of your survey you have been speaking about virgin soils where man had meddled very little, and I am afraid that many of us are concerned with predicting the fertility of soils that have been very much man-altered. For instance, in many of the areas which are very arid you find man putting water on them, as irrigation water. He will also apply fertilisers. In no instance did I notice that man was one of the factors in your considerations and I think this is very important. Would you tell me what are your thoughts on this matter?

[1] *Jenny* - pag. 36

JENNY

We have selected these areas specifically from the viewpoint that as far as we know man had had no effect on it, at least white man. Now, I grant you if we bring in white man and modify the whole system, the situation becomes more complicated, but I thought it was good enough to try to see how the integration of the variables behaves in as simple a system as possible. Accordingly, I must admit that I cannot answer your question at this time.

FLAIG

Dr. JENNY, I followed your paper with great interest because we also tried to find out with statistical methods according to SNEATH the similarities of different humic fractions which are isolated from different soils, and to find out how near are the chemical and physical properties of the same type of different humic substances of different origin. I would like to ask you, did you only determine nitrogen and carbon content of soil organic matter, or have you also investigated parts of your soil organic matter with other methods?

JENNY

This is an embarrassing question and I must say we have not, and not because we underestimate the value of additional studies but simply for the physical impossibility of doing all the detailed field work and the laboratory work and the mathematical work and all that. It is just too big a project for us to bring in at this time your humus science too.

BROADBENT

I noted with interest that you find a much better correlation of nitrogen content with clay minerals than of carbon. In this

connection, have you determined the quantity of ammonium fixed in the clay minerals of these soils?

JENNY

Naturally, when the results emerged, we were wondering what was the role of the clay minerals in the nitrogen correlations. All I can say is that we followed STEVENSON's technique to get all the nitrogen by prolonged Kjeldahl analysis. We have not explored this problem further.

ALEXANDER

If you look at a region in which the plant community has developed in isolation, for example, an island ecosystem, and contrast this with an ecosystem which has a comparable temperature, precipitation, altitude etc., what effects does this isolated plant community have on the levels of soil organic matter or nitrogen? You show no floristic effect, but there is a good chance of introduction of plants in the California region. This does not occur in some isolated ecosystems.

JENNY

Depending on the flora I would expect a vegetation effect, particularly on the carbon/nitrogen ratio and on certain aspects of acidity. Now, you are asking me to compare the isolated island with what?

ALEXANDER

If one examines two plant communities each developing in different conditions, for example an island, or consider an Australian

[1] *Jenny* - pag. 38

locality, a South-East Asian region, or a Western hemisphere environment, does one have the same general results when temperature and precipitation are constant and the variable is the flora?

JENNY

I would, from the studies we have made here, expect a significant differentiation in many soil properties especially in organic matter.

HENIN

Le Docteur JENNY dans son très bel exposé nous a dit, si j'ai bien compris, que pratiquement toute la pédogenèse se trouvait gouvernée par les facteurs climatiques actuels et il a dit qu'il pensait que le problème des sols fossiles représentait peu d'importance étant donnée la vitesse avec laquelle les équilibres pouvaient s'établir. Vous êtes parti d'une région où, si j'ai bien compris, vous aviez comme matériel originel des roches éruptives. Si vous partez d'autres matériaux, par exemple ayant une certaine teneur en argile à l'origine, le fait même que vous trouviez une correlation très étroite entre la teneur en argile et la composition de votre matière organique, devrait amener une évolution différente pour le même facteur climatique. D'autre part, l'évolution que vous pouvez observer pour l'argile est possible parce que vous avez certaines réserves d'élements minéraux, par exemple de silice, pour former un minéral de type vermiculite ou montmorillonite mais si vous partez d'une roche plus pauvre en silice, par exemple une formation sédimentaire avec de la kaolinite, est ce que dans les conditions vous n'auriez pas d'autres formes d'évolution; autrement dit, est ce qu'on peut vraiment généraliser pour tous les types pédologiques ou est ce qu'il ne faut pas prendre, pour chaque type de materiel originel une séquence propre où le facteur climatique trouverait une expression plus spécifique. Est-ce que c'est clair?

[1] *Jenny* - pag. 39

JENNY

Oui, Monsieur.

JENNY

I should like to try to summarise the question in English. You think, Prof. HENIN, that I over-emphasized the climatic variable in clay formation as compared to the composition of the rock? I think I gave a wrong impression. We have always used two parent materials. In both of them the clay mineralogy is a function of climate, but also, as I tried to show, each parent material has its own sequence of clay minerals. For example, there is more montmorillonite in soils from basic igneous rocks and more iron oxide and gibbsite. I think the analysis has shown the importance of the clay mineralogy and the climate for explaining the variability of nitrogen, and the former includes the variation in the parent rocks. I fully concur that soil-climate relations will be different for different types of parent materials.